

AN APPLICATION OF THE MIXED EFFECTS TREND VECTOR MODELS TO THE ANALYSIS OF ASYMMETRIC SQUARE CONTINGENCY TABLES WITH AUXILIARY VARIABLES

Mark de Rooij*

We propose to use the mixed effect trend vector model for modeling of repeated multinomial choice data in the form of a square contingency table. Such data often shows asymmetries where more people change from category a to b than the other way around. In many cases an investigator has, besides the actual choices of the participants, auxiliary variables that pertain to the subjects under study. Most methodologies for asymmetric data do not take into account such variables. We will show how to incorporate these auxiliary variables into the mixed effects trend vector model and how they can be used to study differential change. The models are illustrated in detail with data from the Dutch parliamentary election studies 2006.

1. Introduction

The analysis of asymmetric data has received quite some interest from statisticians. Main examples include the work of Gower (1977), Constantine and Gower (1978), Okada and Imaizumi (1997) and Heiser (Zielman and Heiser, 1993; De Rooij and Heiser, 2000). In these works graphical displays are proposed where the asymmetry is represented through an area, vector or radius in an otherwise symmetric representation. All these methods focus on a simple asymmetric table.

The empirical examples of asymmetric tables in the above mentioned papers often are tables of counts dealing with measurements of participants on two occasions. Another example of such data is given in Table 1 and will be discussed in more detail shortly. For the analysis of such a table by multidimensional scaling procedures the frequencies have to be transformed to dissimilarities. Such a transformation is often performed in a pre-processing step. The disadvantage of such pre-processing steps is that the influence of this transformation cannot be found back in any goodness of fit statistics. That is, the pre-processing is not a part of the model and as such the influence on the outcome of an analysis can not be assessed. Different authors use different transformations: Zielman and Heiser (1993), for example, apply a gravity model which amounts to first dividing the raw frequencies by their row and column totals; second, inverting these standardized frequencies; and third, take the square root of the latter quantities. Okada and Imaizumi (1997) rescale a square table of frequencies with a constant c_j so that the sum of row j plus the sum of column j elements of the rescaled table is equal to the mean sum of row plus column elements. The latter are then used as similarities in a non-metric multidimensional scaling procedure using a monotone transformation. Many other ways of transforming frequencies to-

Key Words and Phrases: biplots, categorical data, Gauss-Hermite quadrature, multilevel model, multidimensional scaling

* Leiden University, Psychological Institute, Methodology and Statistics Unit

wards dissimilarities can be thought of. More importantly, in a practical data analysis situation a researcher might try several ones in order to find a result that best suites him/her.

Although the above mentioned procedures provide neat graphical displays that show the asymmetry none of the procedures is capable of dealing with external variables. In practical data analysis situations the investigator often has auxiliary or supplementary (concomitant, explanatory) variables pertaining to the participants under study. In such a case he or she wants to take into account these variables in the statistical analysis.

In this paper we will use a principled way for the analysis of asymmetric change data where auxiliary variables are available pertaining to the subjects under study and where the transformation is part of the model. In order to do so, we will use a methodology recently proposed by De Rooij and Schouteden (2011), called the mixed effects trend vector model. This methodology combines ideas from generalized linear mixed models and external multidimensional unfolding. This line of research origins in the work of Takane, Bozdogan and Shibayama (1987) and was further worked out by De Rooij (2009a,b); De Rooij and Schouteden (2011).

Before we introduce the model we will outline our application in more detail. In section 3 we revisit the mixed effects trend vector model, provide a detailed description for our application, discuss interpretational issues, and show how the model might be estimated using SAS software. In Section 4 we discuss the results of the mixed effects trend vector model on our application and we conclude in Section 5 with some discussion.

2. Transitions in voting behavior

Consider as an example data from the Dutch Parliamentary Election Studies 2006. Table 1 shows the cross classification of 1569 Dutch inhabitants with their political vote in 2003 against the vote in 2006. Longitudinal categorical data, like these, are often collected in social, medical, geographical and other areas of science and naturally give rise to asymmetries since some people change over time and often changes into one direction are more frequent than changes in the opposite direction. Table 1 shows the distribution of the votes at the two time points and the transitions between seven political parties in the Netherlands: The Christian democratic party (CDA), the labor party (PvdA), the conservative liberals (VVD), the green left party (GL), the progressive liberals (D66), the Socialists party (SP), and the Christian Union (CU).

Looking at the data considerable changes can be seen. Few people leave the CU, while many participants turn towards the CU. A similar pattern can be seen for the SP. Contrarily, D66 loses many votes. Two further noteworthy asymmetries can be found in the 76 participants that voted VVD in 2003 and CDA in 2006 compared to the 31 participants that switch in the other direction and the 111 participants that voted PvdA in 2003 and SP in 2006 compared to 14 that switch the other way around.

For each of the 1569 participants we do not only have information on the two

Table 1: Cross classification of 1569 subjects' vote in 2003 and 2006.

		2006							Total
		CDA	PvdA	VVD	GL	SP	D66	CU	
2003	CDA	365	18	31	2	35	3	17	471
	PvdA	15	309	9	12	111	5	4	465
	VVD	76	8	186	1	11	4	3	291
	GL	4	8	1	46	25	1	5	90
	SP	6	14	0	9	91	0	2	122
	D66	7	14	16	11	16	22	3	89
	CU	1	1	0	0	1	0	38	41
	Total	474	374	243	81	290	35	72	1569

choices but also measurements on six auxiliary variables all measured in 2003. These variables are formulated as follows

- *Income*: Some people think that the differences in incomes in our country should be increased. Others think that they should be decreased. Where would you place yourself on a line from 1 to 7, where 1 means differences in income should be increased and 7 means that differences in income should be decreased?
- *Asylum*: Some people think that the Netherlands should allow more asylum seekers to enter. Others think that the Netherlands should sent asylum seekers, who are already staying here, back to their country of origin. Where would you place yourself on a line from 1 to 7, where 1 means more asylum seekers should be allowed to enter and 7 means asylum seekers should be sent back?
- *Crime*: People think differently about the way the government fights crime. Where would you place yourself on a line from 1 to 7, where at the beginning of the line (1) people are positioned that think the government is acting too tough on crime and at the end of the line (7) people are positioned that think the government should be tougher on crime?
- *Nuclear*: Some people think that nuclear power plants are the solution to a shortage of energy in the future. Others think nuclear power plants shouldn't be built, because the dangers are too great. Where would you place yourself on a line from 1 to 7, where 1 means nuclear power plants should be built quickly and 7 means that they shouldn't be built?
- *Foreign*: In the Netherlands some think that foreigners should be able to live in the Netherlands while preserving their own culture. Others think that they should fully adapt to Dutch culture. Where would you place yourself on a line from 1 to 7, where 1 means that foreigners can preserve their own culture and 7 means that they should fully adapt?
- *Europe*: Some people think that the European unification should go further. Others think that the European unification has already gone too far. Where would you place yourself on a line from 1 to 7, where 1 means that the European unification should go even further and 7 means that the unification has already gone

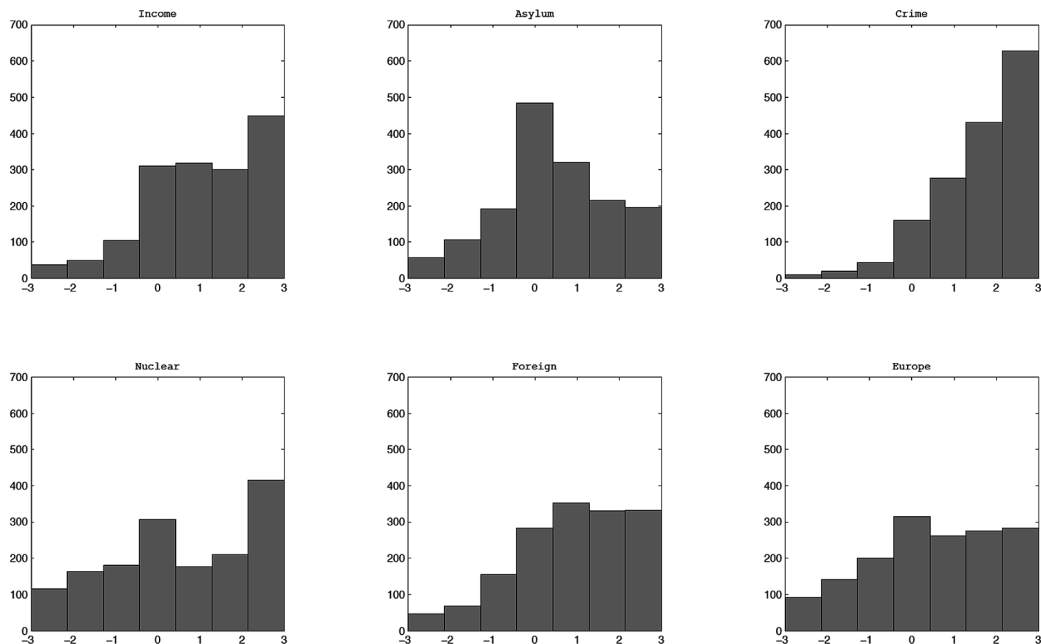


Figure 1: Histograms of the six auxiliary variables

Table 2: Correlations among the auxiliary variables Income (I), Asylum (A), Crime (C), Nuclear (N), Foreign (F), and Europe (E).

	I	A	C	N	F	E
I	1.00	-0.08	-0.01	0.28	-0.09	0.13
A	-0.08	1.00	0.36	-0.08	0.47	0.24
C	-0.01	0.36	1.00	-0.02	0.43	0.23
N	0.28	-0.08	-0.02	1.00	-0.05	0.13
F	-0.09	0.47	0.43	-0.05	1.00	0.24
E	0.13	0.24	0.23	0.13	0.24	1.00

too far?

All these variables are recoded on a seven point scale which we recoded into a scale from minus three to plus three. These variables will be used as continuous auxiliary variables in the subsequent analyses. In Figure 1 the distributions of the six auxiliary variables are shown and in Table 2 the correlations among the auxiliary variables. It can be seen that all variables are skewed, with more positive responses occurring than negative ones. Furthermore, the three variables Asylum, Crime, and Foreigner have the highest correlations; other correlations are all low.

In this paper we will apply the mixed effects trend vector model, recently proposed by De Rooij and Schouteden (2011), to the Dutch parliamentary election study data where we have measurements on two occasions of the political vote of the subjects at national elections. Moreover, we would like to take into account the six auxiliary vari-

ables for every participant and use these variables in order to study whether change is homogeneous.

Before we start outlining the model some general notation is introduced and the way we have to set up our data matrix. The sample consists of $n = 1569$ subjects and for each subject i ($i = 1, \dots, n$) there are measurements on 2 occasions. Let \mathcal{G}_{it} denote the t -th observation ($t = 1, 2$) for subject i , with $\mathcal{G}_{it} = c$ ($c = 1, \dots, 7$) and response probabilities $\pi_{itc} = P(\mathcal{G}_{it} = c)$. Furthermore let \mathbf{g}_{it} be the corresponding vector $\mathbf{g}_{it} = [g_{it1}, \dots, g_{itC}]^T$ with $g_{itc} = 1$ if subject i at time point t chooses political party c and $g_{itc} = 0$ otherwise. For every subject there are $p = 6$ auxiliary variables x_{ij} , $j = 1, \dots, 6$. The general layout of the data as used in this paper is shown in Table 3, the data is in so-called person-time format where for each subject on a specific time point the data are given on a single row. Such a format is common in generalized linear mixed models.

3. The mixed effects trend vector model revisited

A brief outline is given here of the mixed effect trend vector model as proposed in De Rooij and Schouteden (2011). Afterwards a detailed specification of the models for our application is given, plus interpretational guidelines.

3.1 General model

The mixed effects trend vector model is an application of external (restricted) multidimensional unfolding to repeated categorical choice data where the rows correspond to the subjects at the various time-points and the columns to the choice categories. Therefore, the squared Euclidean distance

$$\delta_{itc} = \sum_{m=1}^M (\eta_{itm} - \gamma_{cm})^2$$

between a position of subject i at time point t with coordinates η_{itm} and a position for category c with coordinates γ_{cm} is inversely related to the probability that this subject at that particular time point chooses category c . Notice in such a set-up that the subject are changing over time, while the categories remain stable. The relationship

Table 3: The data format

Subject	Time	Vote	auxiliary variables					
1	1	\mathcal{G}_{11}	x_{11}	x_{12}	x_{13}	...	x_{16}	
1	2	\mathcal{G}_{12}	x_{11}	x_{12}	x_{13}	...	x_{16}	
2	1	\mathcal{G}_{21}	x_{21}	x_{22}	x_{23}	...	x_{26}	
2	2	\mathcal{G}_{22}	x_{21}	x_{22}	x_{23}	...	x_{26}	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
n	2	\mathcal{G}_{n2}	x_{n1}	x_{n2}	x_{n3}	...	x_{n6}	

between the squared distance and the probability is specified by the Gaussian decay function, that is

$$\pi_{itc} = \frac{\exp(-\delta_{itc})}{\sum_h \exp(-\delta_{ith})}.$$

The coordinates of the position for a subject at a specific time point are a linear function of the auxiliary variables and the time variable collected in the vector \mathbf{x}_{it} . These functions are constructed as in any regression model, i.e.

$$\eta_{itm} = \alpha_m + \mathbf{x}_{it}^T \boldsymbol{\beta}_m. \quad (1)$$

Examples will follow shortly. To deal with the dependency among the responses of a given subject two general methods are available. The first is a Generalized Estimating Equation (GEE) approach in which a set of estimating equations is set up that incorporate the dependencies by a working correlation structure. To further deal with the dependencies the variance covariance matrix of the estimates is adapted using a sandwich estimator. The general theory of this methodology is outlined in Liang and Zeger (1996); for the trend vector model see De Rooij (2009b) and Yu and De Rooij (2009). The second approach, the one that we will adopt here, is to include subject specific, or random, effects into the regression equation that actively model the dependencies among the responses. Given the subject specific effects we assume the responses of a subject are independent, an assumption better known as the local independence assumption and one that is very common in subject specific models. This approach is better known as the generalized linear mixed effects approach, or multilevel approach. In our model, the random effects are incorporated into the regression equation for the position of the subject at a given time point:

$$\eta_{itm} = \alpha_m + \mathbf{x}_{it}^T \boldsymbol{\beta}_m + \mathbf{z}_{it}^T \mathbf{u}_{im}, \quad (2)$$

where \mathbf{z}_{it} is the design vector for the subject specific effects and \mathbf{u}_{im} are the subject specific effects. When $\mathbf{z}_{it} = [1]$, only random intercepts are included into the model, whereas in the case $\mathbf{z}_{it} = [1, T_{it}]$ random intercepts and slopes (time effects) are present in the model. For the subject specific effects we assume a multivariate normal distribution

$$\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

These random effects model the dependencies in the data. The larger the variance of the random effects the larger the association between the measurements of a single subject.

With the auxiliary variables we can try to model the individual differences represented by the subject specific effects. That is, by including the auxiliary variables into the model often the variance of the random effects diminishes. Some individual variation is accounted for by the inclusion of the auxiliary variables.

3.2 Detailed Specification and Interpretation

When we ignore the auxiliary variables the only predictor variable is time. Coding time using a dummy variable with $T_{i1} = 0$ for the first time point and $T_{i2} = 1$ for the second, the random intercept model has linear predictor

$$\eta_{itm} = \alpha_m + u_{0i.m} + T_{it}\beta_m, \quad (3)$$

such that $\mathbf{x}_{it} = [T_{it}]$ and $\mathbf{z}_{it} = [1]$ in equation (2). When the time effect may differ over subjects the random intercept and slopes model results

$$\eta_{itm} = \alpha_m + u_{0i.m} + T_{it}\beta_m + T_{it}u_{1i.m},$$

such that $\mathbf{x}_{it} = [T_{it}]$ and $\mathbf{z}_{it} = [1, T_{it}]^T$ in equation (2). For our analysis with only two time points this latter model is close to saturated, and difficult to fit. Therefore we will not further discuss this model.

Both models can be extended with the auxiliary variables. The random intercept model then becomes

$$\eta_{itm} = \alpha_m + u_{0i.m} + T_{it}\beta_{1m} + I_i\beta_{2m} + A_i\beta_{3m} + C_i\beta_{4m} + N_i\beta_{5m} + F_i\beta_{6m} + E_i\beta_{7m}, \quad (4)$$

where I_i represents the value of the Income variable for subject i , similarly A_i for Asylum, C_i for Crime, N_i for Nuclear, F_i for Foreign, and E_i for Europe. In this case $\mathbf{x}_{it} = [T_{it}, I_i, A_i, C_i, N_i, F_i, E_i]^T$.

Although we do not further use the random slopes model, we can still investigate whether change is homogeneous or not. Therefore, we use interactions between the auxiliary variables and the time variable. For example, to verify whether participants with a high value on the Income variable change in a similar way as those with a low value on the Income variable we formulate the following linear predictor

$$\begin{aligned} \eta_{itm} = & \alpha_m + u_{0i.m} + T_{it}\beta_{1m} + I_i\beta_{2m} + A_i\beta_{3m} + C_i\beta_{4m} \\ & + N_i\beta_{5m} + F_i\beta_{6m} + E_i\beta_{7m} + T_{it}I_i\beta_{8m}. \end{aligned} \quad (5)$$

When this model fits better compared to model 4 (that is, the model described in Equation 4) there is differential change. The effect of time for participants with $I_i = -3$ is for example $\beta_{1m} - 3\beta_{8m}$, for participants with $I_i = 0$ it is β_{1m} , and for participants with $I_1 = 2$ it is $\beta_{1m} + 2\beta_{8m}$. So, using the interaction effect we can test the homogeneous change assumption present in model 4. If the test is rejected, we can look at combinations of parameter estimates to interpret the differential change.

Subjects positions can be obtained by two equivalent methods: completing parallelograms or the vector sum method (Gower and Hand, 1996). Note that to obtain a position of a participant his or hers random intercepts have to be added to obtain the exact location. The random intercepts will be represented by an ellipse giving a 68% region. That is, 68% of the subjects fall within this region, i.e. it is an ellipse with major axes equal to the standard deviations of the random effects. The subject

specific effects together with the linear combination of the auxiliary variables determines the position of a subject. With this position the distances towards each of the categories and thus the probabilities for each of the categories can be determined. An important interpretational tool for categorical data is the odds ratio. Given the mixed effects trend vector model, the odds that a specific subject i at time point t chooses category a instead of b are given by $\exp(\delta_{itb} - \delta_{ita})$: the odds are in favor of the closest category.

3.3 Estimation in SAS

In this section we show how these models can be estimated in the SAS statistical software package. We refer to De Rooij and Schouteden (2011) for a detailed description of the estimation procedure. Given the random effects we assume that the responses are independent multinomial distributed variables. The random effects are integrated out of the likelihood using a Gauss-Hermite quadrature scheme. Since we have random effects in the regression equation for each dimension the number of random effects grows quickly with the number of dimensions. In general only models with up to about five or six random effects are estimable. For the estimation we have to specify the number of quadrature points per random effect.

In De Rooij (2009b) it is shown that the trend vector model is not identified. First of all, distances are invariant to translation and rotation. A third indeterminacy, due to the Gaussian decay function is that a constant might be added for each subjects' squared distance without changing the probabilities, that is

$$\pi_{itc} = \frac{\exp(-\delta_{itc})}{\sum_h \exp(-\delta_{ith})} = \frac{\exp(-\delta_{itc} + s_{it})}{\sum_h \exp(-\delta_{ith} + s_{it})} = \frac{\exp(-\delta_{itc}^*)}{\sum_h \exp(-\delta_{ith}^*)}.$$

To obtain identified solutions we will constrain coordinates of the category points (γ_{cm}). We set $\gamma_{1m} = 0$ to deal with the translational indeterminacy; we set $\gamma_{m,m+1} = 0$ to deal with the rotational indeterminacy; finally, by setting $\gamma_{m+1,m} = 1$ the 'scaling' indeterminacy is dealt with. In summary the coordinate matrix for the class points has the following form

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ \gamma_{31} & 1 \\ \gamma_{41} & \gamma_{42} \\ \vdots & \vdots \\ \gamma_{71} & \gamma_{72} \end{bmatrix}.$$

In all cases, we checked whether the gradients are all approximately zero and the matrix of second derivatives is of full rank.

After the algorithm has converged, it is possible to compute the values of the random effects using *expected a posteriori* (EAP) or empirical Bayes estimation. Again

this involves a number of integrals but like in the optimization procedure these can be approximated using quadrature methods.

This whole procedure can be performed using `Proc NLMIXED` of the SAS software. This is shown in the Appendix.

4. Dutch parliamentary election studies

The Dutch political system is often thought to be two dimensional: a left-right continuum and a progressive-conservative one (Pennings and Keman, 2003; Van Holsteyn and Irwin, 2003). Therefore we will do all analyses in two dimensions.

4.1 Model 1: Pure change

First model 3, the model with random intercepts and only time as predictor variable, is used to analyze the political vote data. The solution is shown in Figure 2. In this graphical representation we show the positions of the political parties. Also decision regions are represented; within a certain region the odds are in favor of the political party belonging to that region.

For each participant it's ideal point is given by the random intercept for the first time points and the random intercept plus the time vector for the second time point. The random intercepts are represented by the ellipse, which gives a 68% region. That is 68% of the subjects fall within this region, i.e. it is an ellipse with major axes equal to the standard deviation. Note that we could also represent each participants intercept, but this would make the representation very cluttered. Since the variance of the intercept is large, we present in the right-hand subfigure a more global view of this model. The left-hand side figure represents more detail in the part of the Euclidean space where the political parties are positioned. A large variance of the random effects means that there is a strong auto correlation among the responses of a single individual. Consider a subject with a very large intercept on each dimension, this intercept then determines in a high extend the choices of the subject at both time points. Compare this participant, for example, with a participant having a large negative intercept on each dimensions. The position of this latter participant will fall in a completely different region of the Euclidean space.

There is an overall trend in the direction of GL, SP and CU and away from CDA, VVD, D66 and PvdA. Due to the long-stretched form of the ellipse many CDA voters turn to CU while many PvdA voters turn to the SP.

4.2 Model 2: Explaining the variance using auxiliary variables

In our first step, model 3 was used to analyze the political vote data. In the second step the six auxiliary variables are added to the model, i.e. model 4 is fitted. The likelihood ratio test for these six variables together equals 468.3 with $df = 12$ giving a significant contribution. The six auxiliary variables explain 76.5% of the variance

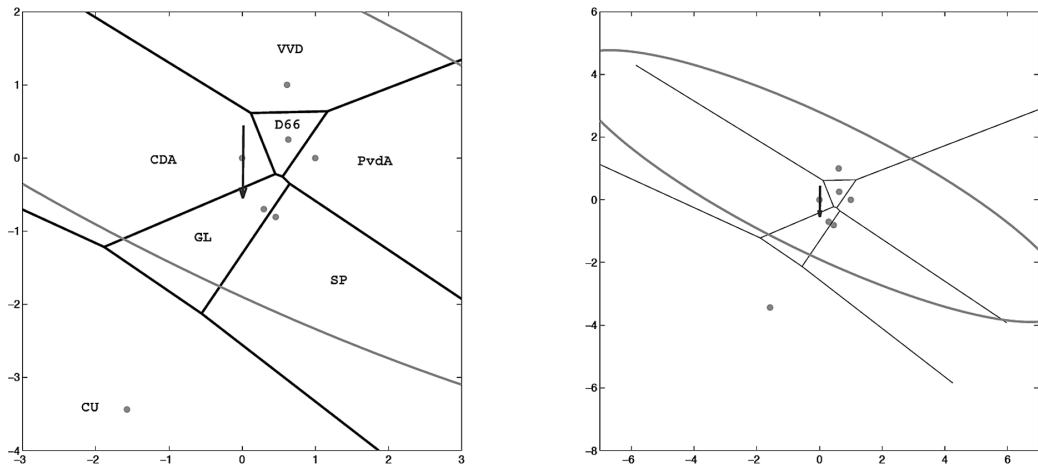


Figure 2: The two dimensional random intercept trend vector model for the Dutch parliamentary election study. The ellipse represents a 68% interval of the random intercepts.

Table 4: Regression weights and test statistics for the auxiliary variables.

Effect	dim	Estimate	SE	LRT
Time	1	0.4508	0.083	29.6
	2	0.0644	0.094	
Income	1	0.7541	0.084	206.3
	2	-0.2787	0.112	
Asylum	1	-0.4656	0.077	52.0
	2	0.0310	0.074	
Crime	1	-0.1490	0.072	12.1
	2	-0.1300	0.051	
Nuclear	1	0.3679	0.052	73.4
	2	0.0261	0.055	
Foreign	1	-0.2525	0.065	27.4
	2	-0.1120	0.052	
Europe	1	0.2612	0.056	28.7
	2	0.0244	0.049	

of the random intercepts on the first dimension and 66.4% of the second dimension. In Table 4 the regression weights with standard errors are given. For every variable we also give the likelihood ratio statistic for the case that that specific variable is excluded from the model. Each variable has a significant contribution.

The solution of this analysis is shown in Figure 3. In this figure the positions of the political parties are given plus regions where the odds to choose them are highest. The trend vector is shown using an arrow and the auxiliary variables are represented using lines with markers for the values -3 , -2 , -1 , 0 , 1 , 2 , and 3 . The labels are given at the positive ($+3$) end of the scale. The only time varying variable is time and

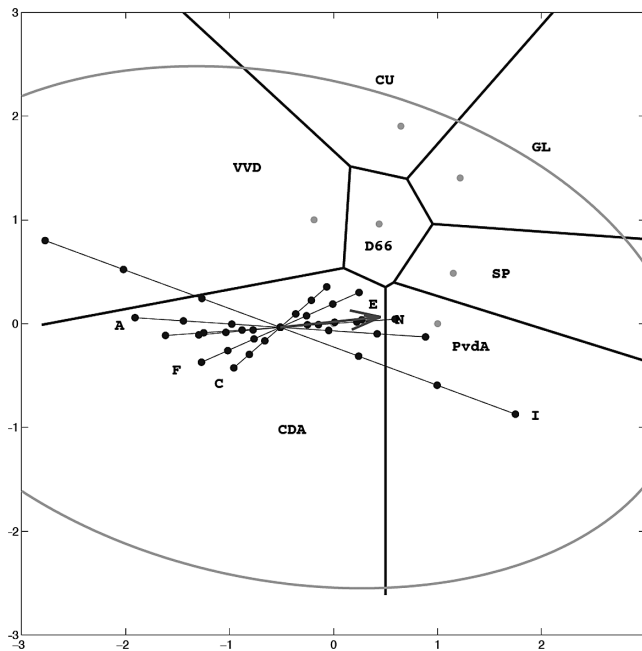


Figure 3: The two dimensional random intercept trend vector model for the Dutch parliamentary election study including auxiliary variables. The variables are labeled at the positive end of the scale: Income (I), Asylum (A), Crime (C), Nuclear (N), Foreign (F), and Europe (E). The ellipse represents a 68% interval of the random intercepts.

since it is in the same direction for all participants, everyone moves into the direction of the trend vector: The change process is homogeneous.

Subjects positions can be obtained by two methods: completing parallelograms or the vector sum method (Gower and Hand, 1996). Note that to obtain a position of a participant his or hers random intercepts also have to be added to obtain the exact location. The random intercepts are represented by the ellipse, which gives a 68% region. Although a large part of the variance of the random intercepts of the first model (equation 3) is explained by the auxiliary variables (the ellipse is much smaller now), still it is quite large representing a strong auto correlation.

Once a subjects position is derived the distances towards each of the political parties can be used as an indication of the probabilities of choosing them. Just from a visual inspection one can derive the rank order of the probabilities. If exact probabilities are needed they can be easily computed from the ideal point and the category positions.

From Figure 3 it can be deduced that

- I*: participants who think that income differences should be decreased ($I = +3$) have a higher probability of voting PvdA or SP, whereas participants that that feel income differences should be increased ($I = -3$) have a higher probability of voting VVD.

- A*: participants who think asylum seekers should be sent back ($A = +3$) have a high probability to vote CDA or VVD whereas participants that would allow more asylum seekers have a higher probability of voting for PvdA, D66, CU, SP, and GL.
- C*: participants who think the government should be tougher on crime ($C = +3$) have a high probability to vote CDA, whereas those who think the government is acting too tough have a higher probability to vote for D66, CU, or GL.
- N*: participants that think nuclear power plants should be built ($N = -3$) have a high probability to vote CDA or VVD, whereas participants who are against new nuclear plants have a higher probability to vote PvdA or SP.
- F*: participants that believe foreigners should completely adapt to the Dutch culture ($F = +3$) have a high probability to vote CDA, whereas those who agree that foreigners can preserve their own culture have a higher probability to vote for D66, SP, or GL.
- E*: participants who think European unification has already gone too far ($E = +3$) have a high probability to vote for PvdA, D66, SP, or GL, whereas those who think European unification should go further have a higher probability to vote for VVD or CDA.

It should be noted that all these directions are conditional relationships, i.e. conditional on the values of the other auxiliary variables and conditional on the random effect.

4.3 Model 3: Testing homogeneous change

As a last step in the analysis process we used model 5, with each of the auxiliary variables in turn, to investigate differential change. The likelihood ratio statistics are 5.2 for Income, 0.2 for Asylum, 0.4 for Crime, 5.6 for Nuclear, 2.2 for Foreign, and 6.1 for Europe, all with two degrees of freedom. Only the latter is significant with an α of $\alpha = 0.05$. The solution is shown in Figure 4. This figure is very much the same as the previous solution (we left out the 68% region for the random intercepts since it is the same as in Figure 3), but now there are multiple trend vectors, for each value of the variable Europe one. We can see that those who think European unification has already gone too far ($E = +3$) tend to change into the direction of the upper right quadrant, while those who think European unification should go further tend to change into the direction of the lower right quadrant.

5. Discussion

We revisited the mixed effect trend vector model for the analysis of longitudinal categorical data and applied it to the analysis of a transition frequency table where for every participant we also have a set of auxiliary variables. Investigators often would like to take into account such variables in the analysis of change. In our empirical

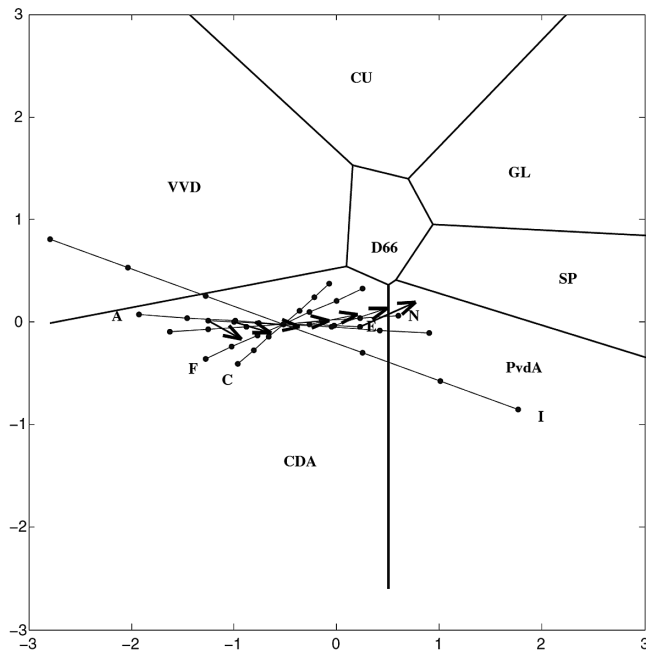


Figure 4: The two dimensional random intercept trend vector model for the Dutch parliamentary election study including auxiliary variables and a differential change process. The variables are labeled at the positive end of the scale: Income (I), Asylum (A), Crime (C), Nuclear (N), Foreign (F), and Europe (E).

example we showed that by taking into account these variables more structure is given to the overall picture. Furthermore, we were able to test for homogeneous change and found a differential time effect.

It can be noted that the solution without auxiliary variables (Figure 2) very much resembles the slide vector model representation as proposed in Zielman and Heiser (1993). We derived this representation from a statistical modeling framework for individual change, whereas Zielman and Heiser derive it from a constrained multi-dimensional unfolding analysis. The slide vector model is built up such that the category points change whereas in the trend vector model the positions of the participants change. In many applications it is more natural to assume that the participants change instead of assuming that the categories of the response variable change. In the empirical example shown, both could in principle change. Political parties change their election program and adapt it to new issues in society; participants change their opinion on certain matters. However, making a model where both (participants as well as political parties) can change is very difficult to identify, since participants and parties can move together into the same direction without changing the probabilities. Further research for such a case is needed.

In the example presented the only time varying auxiliary variable is time. So, the only variable that models the asymmetry, i.e. change, is time. If more time varying predictors would be available a more detailed analysis of the change mechanism could

be given. With time varying predictors it is often useful to decompose them into two new variables: one representing the mean value and one the difference between the actual values and the mean. In such a case the first is a time constant predictor representing an overall preference for a political party; the second one would indicate how change in that predictor is related to change in preference.

The mixed effects trend vector models can be estimated using the `NLMIXED` procedure in `SAS`. This procedure is a general optimization procedure that can take into account normally distributed random effects. The code is given, such that other researchers can easily change it for their own analyses. An important assumption in the mixed effects trend vector model is that the responses are independent given the random effects. Whether this is actually true is hard to examine at this moment. Further investigations are needed that address this issue for clustered multinomial data.

Appendix: SAS code

In Figure 5 SAS code is shown for the estimation of the random intercepts model with both time and the six auxiliary variables as predictors. In the first line `Proc NLMIXED` is called and the data set is specified, `noad` specifies that we will use non-adaptive Gaussian quadrature to integrate out the random effects. In lines 002 till 007 the parameters are defined and starting values are given. Lines 009 and 010 define the linear predictors for the first and second dimensions. To identify the solution some category points have to be fixed, which is done in line 012. Note that these category points are no parameters, they are constants. The squared Euclidean distances are defined in lines 014 to 020. From these distances the probabilities are defined in lines 022 to 029. Line 031 defines the likelihood and line 033 defines the random effect distribution. To ensure that the variances are positive we use the squared terms `s1*s1`, so the absolute value of `s1` gives a standard deviation of the random intercept on dimension one.

Acknowledgements

We acknowledge `DANS` (www.dans.knaw.nl) for making available the data of the Dutch Parliamentary Election Studies. The data for the DPES 2006 were collected by Statistics Netherlands and by Kees Aarts, Henk van der Kolk, Martin Rosema and Martha Brinkman on behalf of the Dutch Electoral Research Foundation (`SKON`). The file obtained has identification number ID: P1719. This research was conducted while the author was sponsored by the Netherlands Organisation for Scientific Research (`NWO`), Innovational Grant, no. 452-06-002.

```

001 proc nlmixed data = dpes noad;
002 PARMs
003 a1 -0.8023
004 a2 -0.2769
005 b11 0.4764
006 ....
007 r12 -0.1276;
008 /* linear predictor*/
009 eta1= u01+ a1 + b11*Time + b21*Income + b31*Asylum + b41*Crime + b51*Nuclear + b61*Foreign + b71*Europe;
010 eta2= u02+ a2 + b12*Time + b22*Income + b32*Asylum + b42*Crime + b52*Nuclear + b62*Foreign + b72*Europe;
011 /* identification restrictions*/
012 z11=0; z12=0; z21=1; z22=0; z32=1;
013 /* squared distances */
014 dist1 = (eta1-z11)*(eta1-z11) + (eta2-z12)*(eta2-z12);
015 dist2 = (eta1-z21)*(eta1-z21) + (eta2-z22)*(eta2-z22);
016 dist3 = (eta1-z31)*(eta1-z31) + (eta2-z32)*(eta2-z32);
017 dist4 = (eta1-z41)*(eta1-z41) + (eta2-z42)*(eta2-z42);
018 dist5 = (eta1-z51)*(eta1-z51) + (eta2-z52)*(eta2-z52);
019 dist6 = (eta1-z61)*(eta1-z61) + (eta2-z62)*(eta2-z62);
020 dist7 = (eta1-z71)*(eta1-z71) + (eta2-z72)*(eta2-z72);
021 /* probabilities */
022 denom = exp(-(dist1)) + exp(-(dist2)) + exp(-(dist3)) + exp(-(dist4)) + exp(-(dist5)) + exp(-(dist6)) + exp(-(dist7));
023 if (vote = 1) then pi = exp(-(dist1)) /denom;
024 else if (vote =2) then pi = exp(-(dist2)) /denom;
025 else if (vote =3) then pi = exp(-(dist3)) /denom;
026 else if (vote =4) then pi = exp(-(dist4)) /denom;
027 else if (vote =5) then pi = exp(-(dist5)) /denom;
028 else if (vote =6) then pi = exp(-(dist6)) /denom;
029 else if (vote =7) then pi = exp(-(dist7)) /denom;
030 /* define loglikelihood */
031 ll=log(pi);
032 model vote ~ general(ll);
033 random u01 u02 normal([0,0], [s1*s1,r12*s1*s2,s2*s2]) subject = id;
034 run;

```

Figure 5: SAS NLMIXED code for estimation of the mixed effects trend vector model with auxiliary variables.

REFERENCES

- Constantine, A.G. and Gower, J.C. (1978). Graphical representation of asymmetric matrices, *Journal of the Royal Statistical Society, series C*, **27**, 297–304.
- De Rooij, M. (2009a). Ideal point discriminant analysis revisited with an emphasis on visualization. *Psychometrika*, **74** 317–330.
- De Rooij, M. (2009b). Trend vector models for the analysis of change in continuous time for multiple groups. *Computational Statistics and Data Analysis*, **53**, 3209–3216.
- De Rooij, M. and Schouteden, M. (2011). The mixed effects trend vector model. *Submitted paper*.
- De Rooij, M. and Heiser, W.J. (2000). Triadic distance models for the analysis of asymmetric three-way proximity data. *British Journal of Mathematical and Statistical Psychology*, **53**, 99–119.
- Gower, J.C. (1977). The analysis of asymmetry and orthogonality. In: BARRA, J.R., BRODEAU, F., ROMER, G., and VAN CUTSEN, B. (Eds.) *Recent developments in statistics*, pp.109–223. Amsterdam: North-Holland.
- Gower, J. C. and Hand, D. J (1996). *Biplots*. London: Chapman and Hall.
- Liang, K-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.
- Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer: New York.
- Okada, A. and Imaizumi, T. (1997). Asymmetric multidimensional scaling of two-mode three-way proximities. *Journal of Classification*, **14**, 195–224.
- Pennings, P. and Keman, H. (2003). The Dutch parliamentary election studies in 2002 and 2003: The rise and decline of the Fortuyn movement. *Acta Politica*, **38**, 51–68.
- Takane, Y., Bozdogan, H., and Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, **52** 371–392.
- Van Holsteyn, J. J. M. and Irwin, G. A (2003). Never a dull moment: Pim Fortuyn and the Dutch parliamentary election of 2002. *West European Politics*, **26**, 41–66.
- Yu, H.-T. and De Rooij, M. (2009). Model selection for the trend vector model. *Submitted paper*.
- Zielman, B. and Heiser, W.J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika*, **58**, 101–114.

(Received February 4 2011, Revised July 15 2011)